

# Linking Asteroid Detections from the Large Synoptic Survey Telescope

Steven R. Chesley  
Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, CA 91109  
steve.chesley@jpl.nasa.gov

Peter Veres  
Minor Planet Center  
Harvard-Smithsonian Center for Astrophysics  
Cambridge, Massachusetts, USA  
peter.veres@cfa.harvard.edu

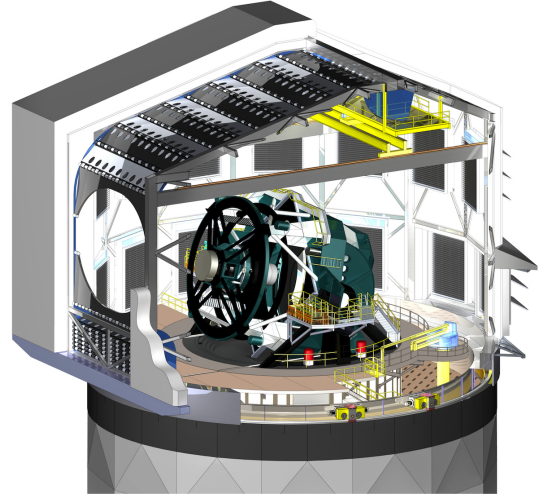
**Abstract**—We have conducted a detailed simulation of the Large Synoptic Survey Telescope (LSST) in order to understand the system’s ability to link detections of asteroids within and across nights in order to populate a catalog of asteroid orbits. We show that LSST, using its baseline survey cadence, should be able to successfully link and catalog asteroids. In our simulation of a single monthly observing cycle, LSST produced 66 million candidate detections of main belt asteroids (MBAs) and near-Earth objects (NEOs), of which 77% were spurious detections related to detector noise or image processing. Using the Moving Object Processing System (MOPS, Denneau et al. 2013, PASP 125, p. 357), we were able to assemble single-night “tracklets” with negligible losses, but a purity of only 43%. The next stage of linking led to three-night orbits with data sets no more than 12 days in length, and it is at this stage that the false detections are readily removed from the data stream. Main-belt linkages were essentially complete and 99.8% pure. Similarly, only 0.02% of linked detections involving NEOs were spurious. On the other hand, NEO linking was 93.6% complete, indicating that 6.4% of potentially findable NEOs were not successfully linked. We believe that this rate can be improved with careful tuning of the MOPS linking algorithms. The NEO catalog was affected by main-belt confusion so that mis-linked MBAs appeared as NEOs, and many correctly linked MBAs were consistent with NEO orbits. We show that these cases arise primarily from MBAs detected at lower solar elongations and we postulate that this is an artifact of a one month simulation that will be readily resolved by surveying over many months.

## TABLE OF CONTENTS

1	INTRODUCTION .....	1
2	LSST MODELING INPUTS .....	2
3	THE LSST DETECTION STREAM .....	3
4	LINKING DETECTIONS TO FORM ORBITS .....	5
5	DISCUSSION .....	11
	ACKNOWLEDGMENTS .....	11
	REFERENCES .....	11
	BIOGRAPHY .....	11

## 1. INTRODUCTION

The Large Synoptic Survey Telescope (LSST) is an ambitious project that has the potential to make significant contributions to Near-Earth Object (NEO) search efforts. LSST is jointly funded by the National Science Foundation and the Department of Energy, with significant enabling contributions from private donors. Construction is well underway and major optical elements are complete. Figure 1 depicts the telescope and dome design in cutaway. LSST first light is set for 2020, followed by two years of commissioning. Once regular survey operations begin in 2022, LSST will systematically



**Figure 1.** A computer rendering of the baseline design of the LSST dome with a cutaway to show the telescope within. Image Credit: LSST.

survey the observable sky over a ten-year period from its site on Cerro Pachon, Chile.

With an 8.4 m aperture (6.7 m effective), 9.6 square degree field of view, and a 3.2-Gigapixel camera, LSST has the potential to become the world’s most capable asteroid survey instrument. LSST will be able to cover over 6000 square degrees of sky per clear night with single visit exposures of 30 s, reaching a faint limit of 24.5 mag in the r band [1]. The survey’s search cadence, however, is a critical factor for NEO discovery performance, and there are multiple science drivers whose different cadence needs are being discussed and will eventually be balanced to shape the final survey strategy.

We have conducted a study to examine the NEO search performance of various LSST search strategies, paying particular attention to the challenges of linking large numbers of asteroid detections in the presence of false detections. The full report is available online [2]; however, in this paper we summarize the linking performance aspect of the complete study. Our approach was to derive lists of synthetic detections for a baseline the LSST survey, based on an assumed model for the populations of solar system objects from the main asteroid belt inwards to the near-Earth population. These detection lists are combined with false detection lists that model both random noise and non-random artifacts resulting from image differencing algorithms. These voluminous detection lists are fed to the Moving Object Processing System (MOPS) [3], which attempts to link the synthetic detections

correctly without becoming confused or overwhelmed by the false detections.

The LSST baseline survey cadence relies primarily on single night pairs of detections, with roughly 30 minutes between the elements of a detection pair. These pairs form what are known in MOPS parlance as *tracklets*, and sets of tracklets are linked across multiple nights to form *tracks*, which can then be sent to the final step of orbit determination. The strategy of using pairs is an aggressive and potentially fragile approach, but theoretically represents the most productive NEO search with the minimum impact on other LSST science drivers. If linking cannot be accomplished successfully for pairs then the LSST project may have to consider alternate observing strategies.

## 2. LSST MODELING INPUTS

The two major questions addressed by our study can be informally stated as “If linking is successful then what fraction of NEOs will LSST discover?” and “Will MOPS be able to successfully link transient detections to form asteroid orbits?” The first question was addressed by quantifying the completeness of the NEO catalog produced by LSST, i.e., the fraction of solar system objects in a given population and size range that LSST would find. Our primary metric for LSST NEO performance was the integral NEO discovery completeness for absolute magnitude  $H < 22$ . These performance results from our study are presented in [2] and [4].

The second question, which is our focus here, asks whether MOPS will work and revolves on the risk of confusion and an ensuing combinatoric crisis in computation. Indeed, the central challenge for the LSST NEO survey is the linking problem, where putative detections of individual moving objects are combined, first within each night, and then across multiple nights, thereby confirming with high confidence that a moving object has been detected and allowing the associated orbit to be computed and cataloged.

Both of these questions required a high-fidelity asteroid detection model to reach an answer, but the modeling approach was markedly different between the two. For the linkage problem we must test MOPS in the presence of confusion due to NEOs, Main-Belt Asteroids (MBAs) and false detections. Thus we must generate full-density detection lists and then feed these lists to the linking engine. The number of detections entering the linking pipeline must match the expected data load of the real LSST, and subtle details in the detection model are less important than assuring the anticipated volume of detections. Thus MOPS testing entails the most computational stressing part of the study, but fortunately these full-density simulations need only take place over a short time period, e.g., a one month-long observing cycle, to understand the MOPS performance. This means that only a tiny fraction of the 10-year survey need be simulated to understand whether MOPS will perform successfully.

In contrast, to obtain the end-of-survey NEO completeness we ran the entire 10-year survey. However, we were able to assume that MOPS would be tested elsewhere, and so we did not need a full-density detection list, nor did we even need to run MOPS. In fact, only the NEO population needed to be included in the simulation; noise and MBAs were neglected. Moreover, a sampling of only a few thousand NEOs was adequate to answer the question. But while the computational load for completeness testing was more manageable, the

fidelity of the detection model was crucial.

With this background in mind, we now turn to a discussion of the key simulation elements that form the framework of the study.

### Operational Simulator

LSST’s Operational Simulator (OpSim) tool [5] uses project scheduling tools to compute all of the field pointings and ancillary information for a full-length, high-fidelity survey, comprising  $\sim 2.5$  million individual field visits over ten years. OpSim models include realistic weather, seeing, sky background noise, etc. For our purposes, the essential OpSim output is a field-by-field listing of the pointings, camera rotation angles, filter selections and the SNR=5 limiting magnitudes  $m_5$ . We used the outputs of OpSim runs as inputs to our study, and our focus for the linking tests was the LSST baseline OpSim run at the time, which was designated `enigma_1189`.

### Solar System Model

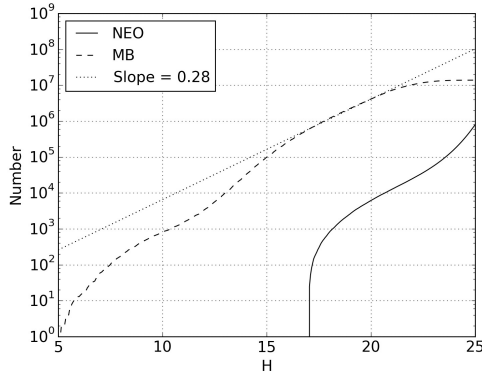
We utilized the Granvik model of the NEO population [6], which contained 801,959 Keplerian orbits with absolute magnitude down to  $H < 25$ . The distribution of its orbital elements is roughly similar to earlier work by Bottke et al. [7]; however, the Granvik population is size-dependent and its size-frequency distribution covers the  $H > 22$  space better than the previous work which underestimated the count. (See Figure 2.) Our NEO population is artificially deficient of large NEO with  $H < 17$ ; however, these are believed to be essentially all discovered and there are only about 500 of them, thus they are a negligible fraction of other detections in an LSST field of view. Initially, we also used the earlier Bottke NEO model [7] that also contains objects down to  $H < 25$ ; however, its total number is only 268,896 and is thus deficient in small objects, particularly for  $H > 22$ .

MBAs will dominate the number density of moving objects in the LSST field of view, and they represent a source of background noise and possible confusion for NEO identification. We used the Grav model [8] of the main-belt population (see Figure 2). This population contains 13,883,361 orbits and is the most robust population model available to date.

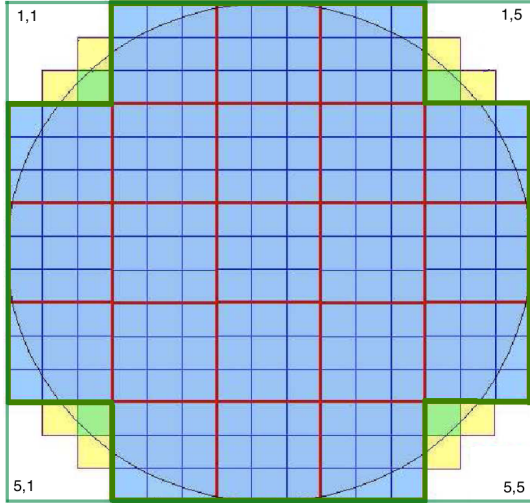
In the Grav MBA model, the cumulative distribution slope is equal to  $\alpha = 0.28 \pm 0.01$  for  $H$  between 16 and 20. However, the population was created for a Pan-STARRS4-like survey with a limiting magnitude of  $m_V = 24.5$ , and so it is truncated to remove MBAs that are fainter than  $m_V = 24.5$  when at perihelion and at opposition. This truncation results in an artificial break, seen in Figure 2, in the Grav population size-frequency distribution at  $H \simeq 21$ . We verified that this artificial break does not have a meaningful effect on the sky-plane densities of detectable MBAs.

### Focal Plane Model

An accurate model of the LSST focal plane is required as a first step in developing high-fidelity detection lists. Figure 3 is a schematic diagram of the LSST focal plane, which consists of 21 CCD *rafts* with each raft comprising a  $3 \times 3$  array of  $4k \times 4k$  CCDs. Thus there are a total of  $21 \times 9 = 189$  CCDs. There are also guide and wavefront sensor CCDs in the corners of the focal plane, but we assume they are not useable. Our modeling approach for the focal plane leverages the existing MOPS formulation, which allows either a square or circular field and allows masking of square sub-regions within the field. We model LSST as a square focal plane with



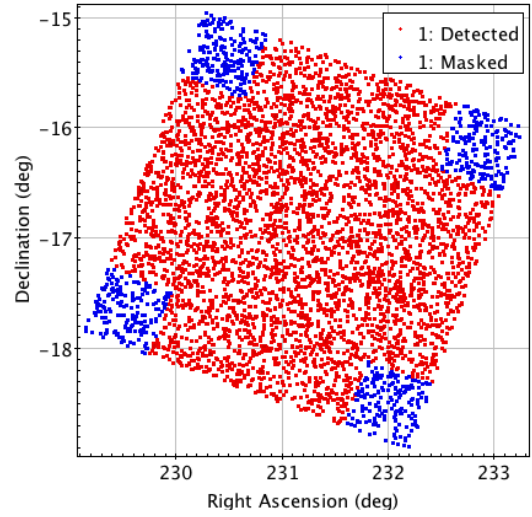
**Figure 2.** Comparison of MBA and NEO size-frequency distributions, where the model MBA slope change at  $H \sim 21$  is an artifact of designing a population of Pan-STARRS4 accessible MBAs.



**Figure 3.** Schematic of the LSST focal plane. The heavy green lines indicate the boundaries of the detection area. The red squares represent the CCD rafts and the small blue squares represent individual CCDs

a  $5 \times 5$  array of rafts and mask the four corner rafts. The LSST focal plane will take different orientations and so we rotate this partially masked focal plane as shown in Fig. 4.

We also need to account for the raft gaps and the smaller chip gaps within each raft. While the exact dimensional specifications for LSST focal plane are not yet finalized, our approach assumes each raft occupies a focal plane area of  $(2500 \text{ arcsec})^2$  and within each raft area the CCD area is  $(2400 \text{ arcsec})^2$ , which is in accordance with the current baseline focal plane design. With these assumptions we find a focal plane *fill factor* of  $F_{\text{fill}} = 92\%$ , which is the fraction of focal plane area that is covered with active CCD pixels. Our current study approach is to statistically under-sample detections according to the fill factor.



**Figure 4.** Depiction of detections and masking for a rotated field. Position Angle (PA) in this example is  $20^\circ$  (angle of rotation from North towards East).

### 3. THE LSST DETECTION STREAM

#### *Synthetic asteroid detections*

We generated synthetic detections for NEO and MBA population models by propagation of the orbits of objects in our Solar System Model to the epochs of the OpSim fields. The propagation used JPL's small body codes with the DE405 planetary ephemerides, where all planets, plus Pluto and the Moon were perturbing bodies. We did not use any asteroids as perturbers. Only detections inside of the rotated field were analyzed and filtered based on the field limiting magnitude and other selected parameters of the detection model.

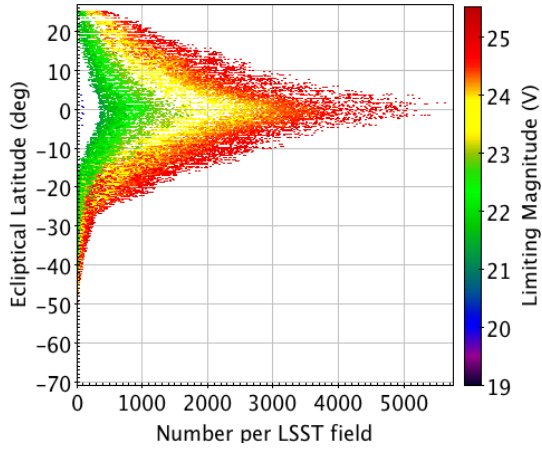
To simulate the observational constraints and limitations of the LSST processing pipeline and CCD effects, we employed a set of filters that determined whether a detection that fulfilled the limiting magnitude was still visible. We included fill factor, vignetting, a smooth fading function (rather than a step function) at the  $\text{SNR}=5$  limiting magnitude  $m_5$ , and trailing losses. The full details of this detection model are discussed in [2] and [4].

Figure 5 shows the spatial density of synthetic NEO and MBA detections in LSST fields as a function of ecliptic latitude and the field limiting magnitude. Here we see that deep fields on the ecliptic plane can have over 5000 asteroids, almost entirely consisting of MBAs. The object density drops off rapidly as fields become shallower and move farther from the ecliptic. Fields more than  $25^\circ$  from the ecliptic show no more than a few hundred detections.

#### *Measurement errors*

Each ephemeris-based position in the field was altered by adding realistic astrometric and photometric errors based on the computed signal-to-noise ratio (SNR). The limiting magnitude of the field  $m_5$  is defined for  $\text{SNR}=5$ . The SNR of a detection is computed from the difference between the computed magnitude  $m$  and  $m_5$  as

$$\text{SNR} = \frac{1}{\sqrt{(0.04 - \gamma) \cdot \chi + \gamma \chi^2}} \quad (1)$$



**Figure 5.** Number of detected MBAs per LSST field as a function of limiting magnitude (V) and ecliptic latitude.

where  $\gamma = 0.038$  and  $\chi = 10^{0.5(m-m_5)}$  [9]. Then, photometric uncertainty is derived as

$$\sigma_m = 2.5 \log_{10} \left( 1 + \frac{1}{\text{SNR}} \right). \quad (2)$$

and the computed  $m$  is combined with an error drawn from a normal distribution with a mean of zero and variance  $\sigma_m^2$ .

We have assumed that LSST astrometry is measured relative to a post-Gaia star catalog and so absolute systematic errors are negligible while relative errors are expected at a floor level of 10 mas. The astrometric error  $\sigma_{\text{astr}}$  for any detection is therefore computed as

$$\sigma_{\text{astr}}^2 = (10 \text{ mas})^2 + \left( \frac{\Theta}{\text{SNR}} \right)^2. \quad (3)$$

Asteroids are moving targets and so, depending on the rate of motion, their shape deviates from a stellar PSF and is in fact a convolution of the motion and the PSF. The faster the object moves, the larger the astrometric error. Therefore, if the trail length  $L > \Theta$ , the seeing term  $\Theta$  in Eq. 3 is replaced by the geometric mean of seeing and trail length as  $\Theta' = \sqrt{\Theta L}$ .

To obtain realistic astrometry, we combine the computed position with an astrometric error term drawn from a normal distribution with a zero mean and variance of  $\sigma_{\text{astr}}^2$ . Figure 6 shows histograms of astrometric uncertainties, in both linear and log-scale. The latter shows that there are two populations of NEA detections, those with high SNR and therefore low uncertainty, around 10 mas, and another centered around 100 mas from low SNR detections, which presumably also includes most of the objects with relatively fast rates of motion. The median astrometric error obtained for NEOs is 47 mas.

#### False detections

The LSST transient detection data stream will include many detections that are not associated with solar system objects, and the objective of linking only detections of real moving objects to form tracks and orbits represents a significant challenge. There are three broad categories of non-solar system transients that are expected from LSST. The first category

of LSST transient detections arise from real astrophysical phenomena (e.g., variable stars, supernovae, etc.) that appear in the same location in multiple instances. Such astrophysical transients will be filtered out of the MOPS input stream by virtue of their stationary appearance and thus will not affect the asteroid linking problem.

The remaining two categories of non-solar system transients consist of spurious detections arising from either random noise or image differencing artifacts, both of which will enter the MOPS input stream. The first source of false detections, from random fluctuations in the sky background and from detector noise, is driven by gaussian statistics at the individual pixel level. The number  $N_{>\eta}$  of these *random* sources above a given signal-to-noise threshold  $\eta$  in the CCD image where Gaussian noise is convolved with a Gaussian PSF follows the formula from [10]

$$N_{>\eta} = \frac{S}{2^{5/2} \pi^{3/2} \sigma_g^2} \eta e^{-\eta^2/2}, \quad (4)$$

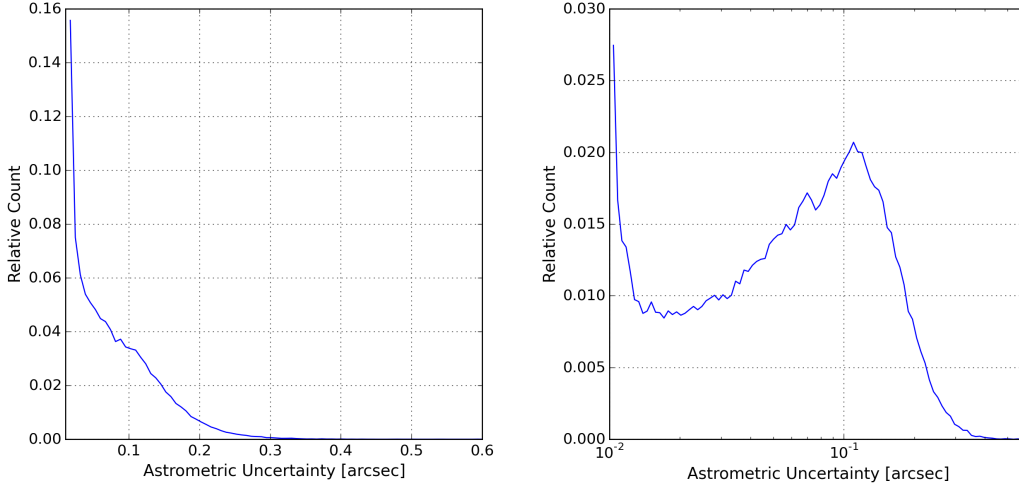
where  $S$  is the total number of pixels in the focal plane array,  $\sigma_g \simeq \Theta/2.35$ , and  $\Theta$  is the FWHM seeing measured in pixels. The number of random false positives depends strongly on the seeing, with about  $\sim 1200$  random false detections per LSST field at  $\Theta = 0.6$  arcsec and only  $\sim 400$  at 1.0 arcsec seeing. The average `enigma_1189` seeing of 0.80 arcsecond leads to 650 random false positives with  $\text{SNR} > 5$  in one LSST image.

We generated random false positives following Equation 4 in random x-y positions in the field. The number of random false positives for a given field was selected from a normal distribution with a mean and variance of  $N$  from equation 4. Then  $\eta$  was randomly assigned according to a normal distribution. Finally, the magnitude can be derived according to  $V = V_{\text{LIM}} - 2.5 \log(\eta/\eta_0)$  where  $V_{\text{LIM}}$  is the  $m_5$  limiting magnitude at  $\eta_0 = 5$ . The number density of random false positives has a strong dependence on SNR; therefore, most of the random noise sources will be near the limiting magnitude (Figure 7).

The second source of false detections comes from difference image artifacts, which arise from differencing a field image with a fiducial image of the static sky that has been derived from a stack of several (or a great many) images of the same field over some time period. This differencing removes stationary objects so that only transient sky phenomena, including moving objects, appear as detections in the difference image. However, registration errors across the field can leave dipole-shaped artifacts in the difference image at the location of a static source. Artifacts may also originate from a poor convolution kernel, variable seeing across the field, stray light in the optical system or reflections in the lenses. Artifacts are often concentrated around bright sources due to saturation or diffraction spikes, and masking around these sources can be an efficient means of substantially reducing the rate of artifacts. Although an improved optical configuration and machine learning can remove many of these false artifacts, some fraction will always remain in the data stream unless unacceptable numbers of real detections are also filtered from the detection stream.

For this work we assumed the estimated density of differencing artifacts derived by Slater et al. [11], who used actual imagery obtained by the Dark Energy Camera (DECam) on Cerro Tololo [12] and processed them with a nascent version of the LSST image processing pipeline. The key result from





**Figure 6.** Distribution of astrometric uncertainties of NEOs - normal scale (left) and logarithmic scale (right).

the Slater et al. report is that “the LSST pipeline is capable of producing a clean sample of difference image detections, at roughly the 200–400 per square degree level.” This is their final result, but our work used a preliminary estimate as the point of departure for our linking simulations. This earlier estimate allowed for roughly 90–380 artifacts per square degree, and we took the geometric mean of this range as the starting point, which leads to  $185/\text{deg}^2$  or 1777 artifacts per LSST field. Slater et al. did find far higher concentrations of artifacts near bright stationary sources, which they eliminated by masking the area around them, thus allowing the reported low artifact density. Following their result, we modeled bright source masking by reducing the effective fill factor by 1%.

To seed the detection list with artifacts, we selected the number of artifacts in each field according to a normal distribution with mean and variance 1777 and distributed them randomly across the field. Thus our artifact rate was roughly  $3\times$  the rate from random noise in typical seeing (Figure 8), and about half of the upper bound derived by Slater et al. [11] from processing actual DECam data.

Based on the Slater et al. report [11], we model that the SNR distribution of differencing artifacts follows  $\propto \text{SNR}^{-2.5}$ . (See Figure 7.) The magnitude of a simulated artifact is then derived according to  $\bar{V} = V_{LIM} - 2.5 \log(\eta/\eta_0)$  where  $V_{LIM}$  is the  $m_5$  limiting magnitude at  $\eta_0 = 5$ . Artifacts have much shallower dependence on  $\eta$ , and therefore tend to be far brighter than random noise sources. Roughly half of modeled artifacts have  $\text{SNR} > 10$ , while virtually none of the random false detections had  $\text{SNR} > 7$ .

#### 4. LINKING DETECTIONS TO FORM ORBITS

The central question for this paper is whether the linking of tracklets into tracks and orbits will prove successful with real LSST data. LSST MOPS will receive full-density lists of detections of moving and transient targets, including NEOs, MBAs and false detections. From these inputs MOPS must create tracklets, tracks and orbits, despite the fact that the data stream is contaminated by potentially large numbers of

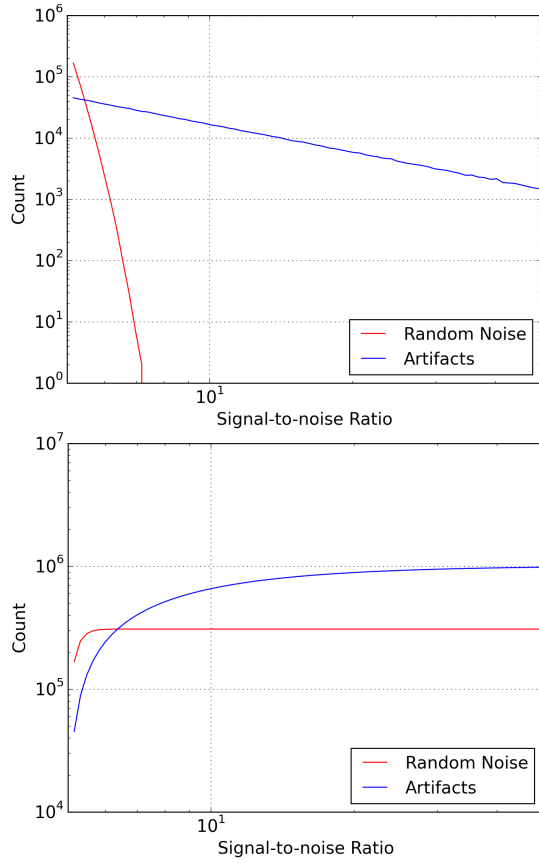
false detections, which leads to high rates of false tracklets. Our simulation synthesized detections in the LSST fields from a full-density NEO model ( $\sim 850,000$  orbits), an MBA model ( $\sim 14$  million orbits) and false detections (both random noise and differencing artifacts). The final detection lists were submitted to the MOPS `makeTracklets` routine, and tracklets were created. Finally, tracklets were submitted to the linking stage, the most challenging step.

##### Tracklets

The list of detections for a given field that has been imaged multiple times in a night is submitted to the `makeTracklets` part of MOPS. A tracklet is created for a detection in the first image if there is a second detection in its vicinity in the second image. The radius of the search circle is defined by the lower and upper velocity thresholds of `makeTracklets`, which were set to  $0.05^\circ/\text{day}$  and  $2.0^\circ/\text{day}$ , respectively, in this study. If there are more possible connections in the circle, in addition to the “CLEAN” tracklet, consisting of detections of one object, then a “MIXED” tracklet consisting of detections of two objects or a “BAD” tracklet that includes a false detection is created as well. Increasing the upper velocity limit increases the search area and thus the number of false tracklets. In some simulations, for velocities of  $1.2\text{--}2.0^\circ/\text{day}$ , we used the information on the trail length to limit the search area for companion detections. At  $1.2^\circ/\text{day}$ , a detection will have a non-PSF shape and its length will be 1.8 times the PSF for the average 0.86 arcsec seeing, and so its length and orientation can be determined. Thus, instead of a large circular search area around trailed detections, smaller regions consistent with the anticipated velocity and direction of the trails are searched, and any matching detections must have a compatible trail length and direction. See Figure 9 for a graphical depiction.

The number of tracklets depends on the density of detections, which can be high (Figure 10). To understand the feasibility of the simulation we gradually increased the number of detections in OC 28. The following steps are also summarized as Cases A-E in Table 1.

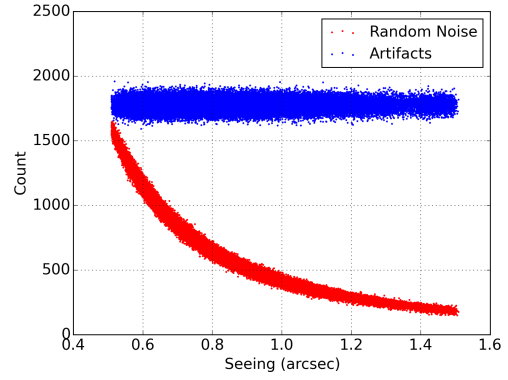
1. Initially, we only used NEO orbits from Bottke’s model (Case



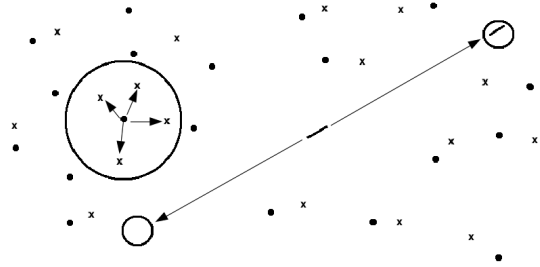
**Figure 7.** Histogram (top) and cumulative distribution (bottom) of random noise and artifacts on one night of LSST survey.

A, Table 1). Switching to Granvik’s NEO model increased the number of detections by 35% and tracklets by 55% (Case B). Because Granvik’s NEO model is more current and has many more objects we used that population in the simulations. At this stage, with only NEO orbits, nearly all tracklets were CLEAN, with only 4 MIXED tracklets (99.97% tracklet purity).

2. Adding the MBA population to Granvik’s NEOs (Case C) increased the number of detections in one month to 15 million and the number of tracklets to 6 million. Most of the tracklets were for MBAs; however, about 17% of tracklets were MIXED, i.e., derived from different objects. The large number of MIXED tracklets was substantially reduced by taking advantage of trail-related velocity information when in the velocity range  $1.2\text{--}2.0^\circ/\text{day}$  (Case D). In this dual velocity mode of makeTracklets,  $1.2^\circ/\text{day}$  is the upper threshold for creating a tracklet by searching in a circle. If the detection is trailed and the trail length implies a velocity  $> 1.2^\circ/\text{day}$ , then its matching pair in the second image must be in a predicted location, based on the time between exposures, and the position and velocity of the first detection (Figure 9). Thus, the number of randomly linked detections in a large circle decreased dramatically. This increased the number of good NEO tracklets by 20% and decreased the number of MIXED tracklets by a factor of 5.
3. The next step added false detections from random noise to the full-density NEO and MBA detection list (Case E). This doubled the number of detections to 30 million, and so the synthetic to false detection ratio was about 1:1. However, the



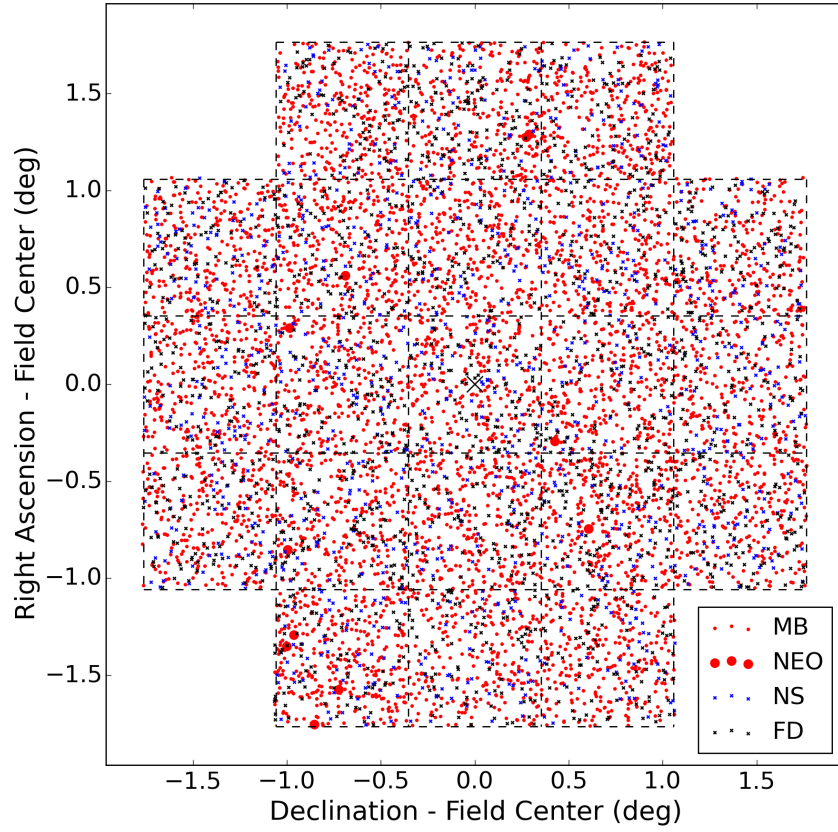
**Figure 8.** Random noise and artifact counts per individual field as a function of seeing during one month of the LSST survey.



**Figure 9.** Schematic diagram for tracklet generation. Dots represent detections from the first image, x signs from the second one. The large circle represents the upper velocity limit for creating tracklets without rate information (up to  $1.2^\circ/\text{day}$ ). Arrows in that circle are all possible tracklets, connecting the first detections with all detections from the second image in the reach. Every detection in the image has such a circle and corresponding set of tracklets. If the detection is faster than  $1.2^\circ/\text{day}$  it will be trailed (on the right), and information on the trail length and orientation can be used to search a smaller area for its counterpart in the second image (in two separate regions because the direction of motion is unknown). The matching detection must also be a trail with similar length and orientation.

number of tracklets only increased from 6 million in case C to 7.5 million in case E. In this scenario tracklets were created up to the  $2^\circ/\text{day}$  limit without the use of velocity information. In addition to 1 million MIXED tracklets, the simulation generated about 700,000 BAD tracklets (i.e., those with both synthetic and false detections) and 600,000 NONSYNTH tracklets consisting solely of false detections.

4. The final, full-density simulation was achieved by also injecting differencing artifacts, which more than doubled again the total number of detections, to 66 million (Case F). Now, over 77% of detections were false, and so the ratio between synthetic and false detections was about 1:3.5. NEOs represent only 0.07% of the detection list. The full-density simulation was challenging for the tracklet stage. Therefore, we used trail-derived velocity information for tracklets created in the velocity range of  $1.2\text{--}2.0^\circ/\text{day}$ . Still, the total number of



**Figure 10.** An example of a high-density LSST field from the `enigma_1189` survey. The depicted field is number 1891 from night 3052, taken in the `r` filter with  $m_5 = 24.79$ , seeing 0.63 arcsec, airmass 1.04, and field center at opposition-centered ecliptic coordinates (Lat., Long.) =  $(2.91^\circ, 1.26^\circ)$ . Thus the field is near opposition in excellent conditions. The various types of detections referenced in the legend are “MB”—main-belt asteroids, “NEO”—near-Earth objects, “NS”—false detections from random noise, and “FD”—false detections arising from image differencing artifacts.

tracklets was very large,  $\sim 11.9$  million. Out of this sample, about 57% of tracklets were somehow erroneous, either including at least one false detection or detections of different objects. This simulation revealed that artifacts related to false positives create the majority of the linking challenge. Though we did not directly test it, the use of trail-related velocity information presumably leads to a dramatic reduction in the false tracklet rate for the full-density simulation.

#### The Linking Process

Automated linking of tracklets is a crucial element of LSST’s NEO discovery pipeline. Without an automated linking stage, the NEO discovery rate would suffer and would rely heavily on follow-up observers, which will be impractical given the faint limit and volume of the LSST detections. The MOPS linking algorithm connects tracklets from three distinct nights into candidate tracks that are subsequently tested through orbit determination. The process consists of the following four distinct steps:

1. *Assemble tracklet list.* The first step collects, for a given field, all of the tracklets from the last  $N$  nights for which the earlier position and velocity project into the destination field. The forward mapping of tracklets is based on linear motion, and acceleration that leads to nonlinear motion is not accounted for. Thus some NEO tracklets may be neglected, especially those very near the Earth with a rapid change in geometry and

observable acceleration.

The combinatorics of linking strongly favor small  $N$ , but the objective of NEO completeness favors large  $N$ , which allows more missed detection opportunities. For LSST,  $N$  usually ranges from 12–20 days, though 30 days has been contemplated as a stretch goal. This work used  $N = 12$  days for linking tests, consistent with our objective of understanding whether linkage could be at all successful in the presence of large numbers of false detections. NEO linkage of nearby objects is not likely to succeed for large  $N$  unless MOPS is extended so that some plane-of-sky acceleration is allowed when assembling the field-by-field tracklet lists. This would likely lead to a modest increase in the NEO discovery rate at the expense of many more false tracklets and increased linking overhead.

2. *Assemble candidate track list.* The second step in linkage generates a list of candidate tracks based on the input tracklets. Generally, there are hundreds of available fields per night, each being processed in parallel. The `linkTracklets` algorithm is based on a kd-tree search [13] that reduces the number of potential tracks to be tested from  $n^2$  to  $n \log n$ , where  $n$  is the number of tracklets available for linking on the given field. This saves a significant amount of computational resources, but the problem remains challenging.
3. *Derive preliminary orbit.* The third step takes the candidate tracks derived by `linkTracklets` and submits them

for Initial Orbit Determination (IOD). MOPS uses Gauss' method to generate potential initial orbits from the astrometry, and for each track the best fitting IOD is selected. Most false tracks were eliminated at this stage with no valid IOD.

4. *Perform differential corrections.* The fourth stage was Orbit Determination (OD), which used JPL OD routines to obtain converged orbits. MOPS filtered out some false tracks at this stage based on rudimentary screening on post-fit residual statistics. As discussed below, MOPS's built-in orbit quality filtering is not strict and is agnostic regarding the expected errors in the astrometry, and thus relatively few false orbits were rejected at this stage. All orbits that passed the MOPS default quality screening were added to the MOPS derived object table, which was the basis for understanding the overall linking performance.

### Linking Performance

Linking tests were conducted on observing cycle 28 of the `enigma_1189` baseline survey, with Granvik's NEO model, MBAs and the full false detection lists (Case F, Table 1). The NEO linking efficiency is defined as the number of unique NEOs present in the post-linking, derived-object catalog divided by the number of unique NEOs with possible 12-day tracks in the detection list. The linking efficiency was 93.6% for  $H < 22$  NEOs and 84.0% for all NEOs (i.e.,  $H < 25$ ). These numbers were lower than the case without the false detections, where we achieved  $> 99\%$  linking efficiency, similar to previous work [13], [3]. The lower efficiency for all NEOs arises from the fact that the vast majority of NEOs were of the smallest diameters, e.g.,  $23 < H < 25$ . Also, smaller objects tend to have faster rates and greater acceleration because they are seen at closer geocentric distances, and they tend to have shorter observability windows. Note that the derived linking efficiency was for a single set of selected kd-tree parameters with a single 8-core workstation. With more powerful computational facilities and a more optimized kd-tree search (possibly on a per-field basis), there is excellent reason to believe that the linking efficiency can be significantly improved.

Many derived NEO orbits stemmed from objects in the MBA input catalog. Table 2 shows the makeup of the 5348 NEO orbits (defined by perihelion distance  $q < 1.3$  au) derived from OC 28 alone. Among these orbits, 2222 originated from CLEAN linkages of actual NEOs, 1896 were CLEAN orbits associated with MBAs and 1230 were erroneous ("Not CLEAN") linkages. Nearly all of the erroneous linkages combined detections of different MBAs to form an NEO orbit; few were contaminated by false detections. At first blush this implies a purity of 77.0% in the NEO catalog, but we describe below why this apparently low accuracy is mostly a manifestation of an ineffective orbit quality screening applied by MOPS. Correct interpretation of the orbits and improved screening increases the accuracy to 96%. In contrast to the NEO orbits, Table 2 reveals that the MBA catalog has 99.8% purity already at this stage, without more refined filtering on orbit quality. Only 6 NEOs appear in the non-NEO orbit catalog, and most of these are borderline cases where  $q \simeq 1.3$  au.

### Orbit Quality Filtering

The large fraction of erroneous linkages that appear in the NEO orbit catalog stem from a weak orbit quality filter implemented by MOPS, which requires the post-fit RMS of astrometric residuals to be less than 0.4 arcsec, a criterion that is too readily met for astrometry with a median error less than 0.05 arcsec. Moreover, because the RMS is not normalized

**Table 1.** Number of detections and tracklets for OC 28 in various simulations. MIXED tracklets include detections from at least two distinct objects, BAD tracklets include detections from both false detections and moving objects, and NONSYNTH tracklets consist entirely of false detections.

Case	NEO	MBA	False Det	Detections				Tracklets			
				Total	%NEO	%MBA	%False	Total	%NEO	%MBA	%MIXED
A	Botke	No	None	36k	100	0	0	11k	100	0	0
B	Granvik	No	None	49k	100	0	0	17k	100	0	0
C	Granvik	Yes	None	15M	0.3	99.7	0	6.2M	0.23	82.8	16.9
D <sup>a</sup>	Granvik	Yes	None	15M	0.3	99.7	0	5.4M	0.31	94.8	0
E	Granvik	Yes	Random only	30M	0.2	50.6	49.2	7.5M	0.19	68.2	14
F <sup>a</sup>	Granvik	Yes	Random + artifacts	66M	0.1	22.6	77.3	12M	0.14	42.7	2.2
										6.1	48.8

<sup>a</sup>Tracklet generation used rate information from 1.2–2.0°/day. Otherwise rate information was ignored over entire range 0.05–2.0°/day.

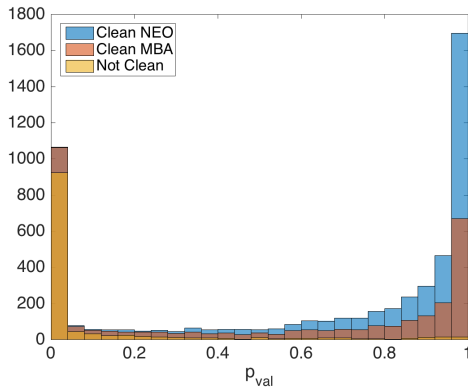


**Table 2.** Accuracy of derived orbits from OC 28. The “Incorrect Class.” column indicates the number of objects for which the source object and the derived object had a different classification based on perihelion distance  $q$ . “Not CLEAN” indicates erroneous linkage of observations from either false detections or multiple objects.

Derived Classification	All	Incorrect Class.	Not CLEAN	Accuracy
NEO ( $q \leq 1.3$ au)	5348	1896 Non-NEO	1230	77.0%
Non-NEO ( $q > 1.3$ au)	765,833	6 NEO	1635	99.8%

by the reported astrometric uncertainty, it fails to take into account the varying quality of astrometry within and between tracklets in a candidate track. The upshot of this approach is that most such erroneous linkages show residuals clearly inconsistent with the astrometric uncertainty, and yet they pass the MOPS quality control test. Rather than modifying MOPS and re-running the simulation, we post-processed the post-fit astrometric residuals, with their associated uncertainties, to derive the sum of squares of the normalized residuals for each orbit in the NEO catalog. This provided the so-called  $\chi^2$  of residuals, from which it is straightforward from classical statistics to calculate the probability  $p_{\text{val}}$  that the fit is valid, which is to say, the likelihood of getting a higher value of  $\chi^2$  by chance. A higher post-fit  $\chi^2$  naturally leads to a lower  $p_{\text{val}}$  because the increased residuals reflect a poorer fit that has a lower probability.

Figure 11 depicts the distribution of  $p_{\text{val}}$  among the 5348 cataloged NEO orbits. The histogram reveals that few erroneous linkages appear for  $p_{\text{val}} > 0.25$  and that few NEOs appear for  $p_{\text{val}} < 0.25$ , thus we selected 25% as the  $p_{\text{val}}$  cutoff for acceptable orbits. This criterion led to rejection of 7% of clean and 87% of not clean orbits. Most of the clean orbits that were filtered out were MBAs mis-classified as NEOs, 14% of which were filtered out. Only 2% of clean NEO orbits were removed by this filter. As tabulated in Table 3, more aggressive  $p_{\text{val}}$  filtering—at the 50% or 90% level—is less effective at removing erroneous linkages, even as the loss of clean NEOs becomes unacceptable. Thus a modest modification of MOPS is necessary to allow a more statistically rigorous orbit quality filtering, but the rudimentary approach described here leads to a 96% purity (3816/3979, see Table 3) in the NEO catalog. In the context of accuracy, the clean MBAs that appear in the NEO orbit catalog are accounted as correctly linked, which is, in fact, the case.



**Figure 11.** Histogram of postfit residual statistics of derived NEO orbits. In most cases, Not CLEAN NEO candidates can be easily distinguished.

The rate of contamination of NEO orbits by false positives is

**Table 3.** The number of cataloged NEO orbits of various classifications for varying values of the  $p_{\text{val}}$  orbit quality filter. Here “Non-NEO” refers to MBAs that appear in the derived NEO catalog with  $q < 1.3$  au.

Classification	0%	$p_{\text{val}}$ cutoff 25%	50%	90%
All	5348	3979	3636	2314
CLEAN	4118	3816	3532	2279
Not CLEAN	1230	163	104	35
w/False Detection	35	3	1	1
CLEAN NEO	2222	2180	2062	1375
CLEAN MBA	1896	1636	1470	904
Not CLEAN NEO	2	0	0	0
Not CLEAN MBA	1228	163	104	35

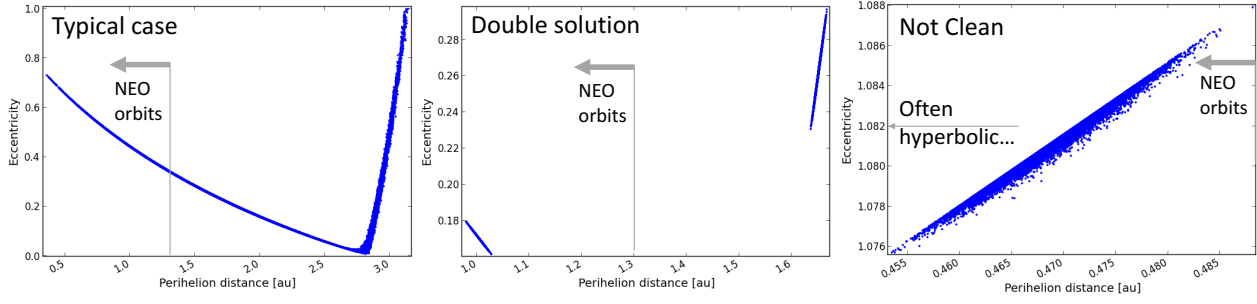
extremely low, despite the large numbers of false positives injected into the detection stream. As shown in Table 4, after filtering at  $p_{\text{val}} > 25\%$ , only 5 false detections appear in the NEO catalog. This can be compared to the total of over 29,000 detections that form the NEO catalog and the 51M false detections polluting the data stream. This result demonstrates that NEOs can be successfully linked with high efficiency and high accuracy when surveying with the baseline LSST cadence, even in the presence of significant numbers of false detections.

**Table 4.** Number of detections of various classifications from OC 28. The total number in the input detection list and the number that were linked into the derived NEO catalog are shown.

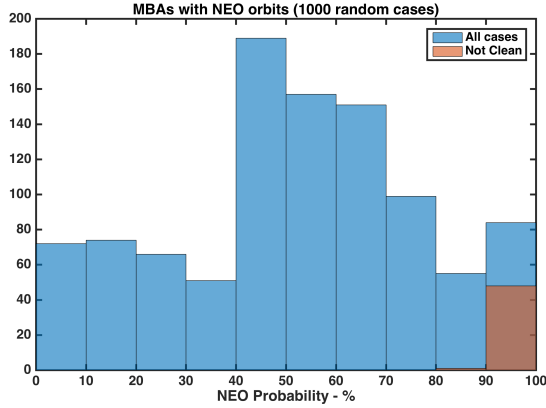
	Total	—Derived NEO Catalog—	
		All	$p_{\text{val}} < 25\%$
Total	65,900,928	39,188	29,288
MBA	14,899,279	20,680	11,868
NEO	48,628	18,446	18,060
False	50,953,021	62	5
% False	77.3%	0.16%	0.02%

#### Confusion from MBAs

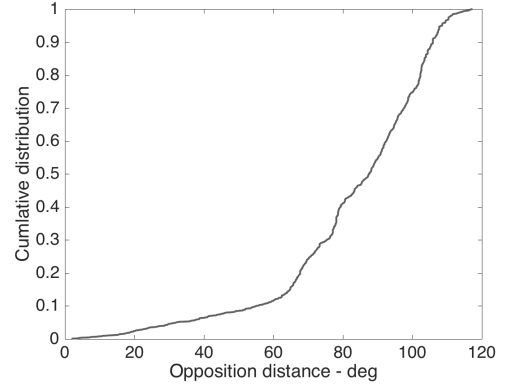
To better understand the issue of the large fraction of NEO orbits stemming from correctly linked non-NEO objects, we used systematic ranging to explore the full orbit determination problem for these cases. Systematic ranging is an orbit estimation technique designed to analyze poorly constrained orbits, typically with only one or a few nights of astrometry, for which the conventional least squares orbit determination can fail due to nonlinearity [14]. We tested hundreds of cases and found that nearly all showed a characteristic “V”-shaped orbital uncertainty pattern in  $e$  vs.  $q$  that allowed both NEO and MBA orbits (left panel, Figure 12). In some cases the “V” shape was broken at the vertex so that there were two distinct orbital solutions (center panel, Figure 12). The systematic ranging technique affords a statistically rigorous estimate of the probability that the track represents an NEO orbit,



**Figure 12.** Examples of typical uncertainty regions for misclassified or erroneous linkages in the derived NEO orbit catalog. The plots depict Monte Carlo samples from systematic ranging that reflect the extent of possible solutions in perihelion distance  $q$  and eccentricity  $e$ . The plots show the typical case of an MBA discovery (left) where the data are compatible with orbits spanning the NEO and MBA orbital regimes. In some of such cases two disjoint solutions are present, one NEO and one MBA (center). Erroneous linkages of two different MBAs often lead to NEO orbits with a small uncertainty, though many such cases are also hyperbolic.



**Figure 13.** Histogram of computed probability that a track derived from MBA tracklets relates to an NEO orbit, as derived from systematic ranging analyses.



**Figure 14.** Cumulative distribution of opposition distance for MBAs that appear in the NEO orbit catalog with  $p_{val} > 25\%$ . The distribution shows that this main-belt confusion is largely limited to detections made far from opposition, i.e., with low solar elongation.

and for these correctly-linked MBAs that appear with NEO orbits, few have high NEO probabilities, reflective of the fact that the data are compatible with the non-NEO (truth) orbits (Figure 13). It is also important to note that most of these MBAs that appear as NEOs are detected far from opposition. Figure 14 shows that only  $\sim 10\%$  of these cases are found within  $60^\circ$  from opposition, and that about half are detected at  $80^\circ$  or farther from opposition. This result is merely reflecting the classical result that orbital ambiguities result from three-night orbits of objects far from opposition. It is an unavoidable feature of observing at low solar elongations, and is generally corrected after a fourth night of data is obtained. However, as described below, the current MOPS configuration does not efficiently attribute a fourth night of data to the already cataloged orbit, and so the ambiguity is often not resolved in our simulations. We note also that this confusion is an artifact of simulating only a single observing cycle. In actual operations, MBAs seen at low solar elongation would eventually move into the opposition region and appear even brighter there. These MBAs would be readily cataloged with their correct orbits because there is little ambiguity in the opposition region, at which point it becomes straightforward to link to the ambiguous orbits arising from near-sun detections.

We also conducted systematic ranging analyses on some of the erroneous linkages leading to NEO orbits, almost all of which were erroneous MBA-MBA linkages, and these revealed a very different characteristic pattern in the  $e$  vs.  $q$  uncertainty space (right panel, Figure 12). The uncertainty region was typically very small, leading to a high computed probability that the orbit is of an NEO (“Not Clean” in Figure 13). In these cases, the uncertainty regions were also elongated and with one side having a sharp cutoff. In many such cases the heliocentric orbits were hyperbolic. This points to a likelihood that more effective screening tests can be developed to eliminate these false MBA-MBA linkages, despite the fact that some pass even strict orbit quality tests. For example, Table 3 shows that even for  $p_{val} > 90\%$  a few dozen erroneous linkages remain in the NEO catalog. However, most of these erroneous MBA-MBA linkages are readily repaired when the individual MBAs are eventually re-observed at other epochs and correctly linked through other tracklets.

## 5. DISCUSSION

We performed a high-fidelity simulation of linking NEO and MBA detections into orbits in a realistic density scenario with false detections and constraints of the LSST survey in one observing cycle. Tracklet generation created false tracklets at a rate of 57% being false. This rate can be larger if one neglects the information on trail length and orientation when creating tracklets. We used this velocity information for the velocity range of 1.2–2.0 deg/day.

On a single-lunation, full-density simulation, with NEOs, MBAs and false detections, we obtained a linking efficiency of 93.6% for  $H < 22$  NEOs with 12-day tracks. Linking efficiency on the full population down to  $H < 25$  was lower. We believe that, with modest revision and tuning of the MOPS linking algorithms and an appropriate allocation of computational resources that this number can be significantly increased, probably to 99% or more.

On the same simulation, the derived NEO catalog was comprised of 96% correct linkages. The remaining 4% of linkages were almost exclusively incorrect MBA-MBA links, most of which should be eliminated over a longer duration simulation. Less than 0.1% of orbits in the derived NEO catalog included false detections.

Some enhancements to MOPS are needed in the linking stage to eliminate duplicate and false orbits. This includes improving the orbit quality filter and tuning of the attribution, precovery<sup>2</sup> and orbit-orbit identification modules. Together with optimization of the kd-tree track search, this would increase the linking efficiency and thus increase the number of cataloged NEOs. The linking efficiency directly affects the discovery completeness as discussed in [4].

## ACKNOWLEDGMENTS

We are grateful for the cooperation of the LSST Project through our consultations with Željko Ivezić, Lynne Jones, and Mario Jurić of the Univ. of Washington. This work was conducted at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration.

## REFERENCES

- [1] Ivezić, Ž., et al., 2014, “LSST: From Science Drivers to Reference Design and Anticipated Data Products,” <http://arxiv.org/abs/0805.2366>
- [2] Chesley, S.R. and Vereš, P., 2017, “Projected Near-Earth Object Discovery Performance of the Large Synoptic Survey Telescope,” JPL Publication 16-11. [https://cneos.jpl.nasa.gov/doc/LSST\\_report\\_2017.html](https://cneos.jpl.nasa.gov/doc/LSST_report_2017.html)
- [3] Denneau, L., et al., 2013. Publications of the Astronomical Society of the Pacific, 125, 357–395.
- [4] Vereš, P. & Chesley, S.R., 2017, Astronomical Journal, submitted
- [5] Delgado, F., Saha, A., Chandrasekharan, S., Cook, K., Petry, C., Ridgway, S. 2014. Society of Photo-Optical In-

strumentation Engineers (SPIE) Conference Series 9150, 915015.

- [6] Granvik, M., Morbidelli, A., Jedicke, R., et al. 2016, Nature, 530, 303
- [7] Bottke, W. F., Morbidelli, A., Jedicke, R., et al. 2002, Icarus, 156, 399
- [8] Grav, T., Jedicke, R., Denneau, L., et al. 2011, Publications of the Astronomical Society of the Pacific, 123, 423
- [9] Ivezić, Z., Tyson, J. A., Axelrod, T., et al. 2009, Bulletin of the American Astronomical Society, 41, 460.03
- [10] Kaiser, N., The Likelihood of Point Sources in Pixelated Images, Pan-STARRS Project Management System, PSDS-002-010-00.
- [11] Slater, C., Jurić, M., Ivezić, Ž., Jones, L., 2016. <http://dmtn-006.lsst.io>
- [12] Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, Astronomical Journal, 150, 150
- [13] Kubica, J., Denneau, L., Jr., Moore, A., Jedicke, R., & Connolly, A. 2007, Astronomical Data Analysis Software and Systems XVI, 376, 395
- [14] Farnocchia, D., Chesley, S. R., & Micheli, M. 2015, Icarus, 258, 18

## BIOGRAPHY



**Steve Chesley** received his B.S. and M.S. degrees in Aerospace Engineering from Texas A&M University in 1986 and 1993, and his Ph.D. in Aerospace Engineering from the University of Texas at Austin in 1998. He is currently a Senior Research Scientist in the Solar System Dynamics Group at NASA’s Jet Propulsion Laboratory, where he led the development of Sentry, JPL’s automated asteroid impact monitoring system. His long-term research focus is on impact hazard assessment and precision orbit determination and ephemeris prediction for small solar system bodies.



**Peter Vereš** received his M.S. degree in Physics and Astronomy in 2006 and Ph.D. degree in Astronomy and Astrophysics in 2010 from the Comenius University in Bratislava, Slovakia. From 2011 to 2014 he worked as a Pan-STARRS1 MOPS Postdoctoral Fellow at University of Hawaii. He was a Caltech Postdoctoral Scholar at NASA’s Jet Propulsion Laboratory from 2015–2017, and now serves as a staff astronomer at the Minor Planet Center. His research activities include NEO survey simulations and search performance, data mining for precovery and follow-up of NEO observations, and physical properties of asteroid populations derived from big data and surveys.

<sup>2</sup>Here “precovery” refers to a search of the MOPS database for tracklets observed previously that did not form a derived object because not enough tracklets were observed at the time. It is similar to attribution of new detections, but operates on past observations.